

XIV Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB 2013)
GT 8: Informação e Tecnologia

Comunicação Oral

CICLO DE VIDA DOS DADOS E O PAPEL DA CIÊNCIA DA INFORMAÇÃO

Ricardo Cesar Gonçalves Santana – UNESP

Resumo

A importância crescente que o acesso a dados vem obtendo em todas as áreas apresenta-se como novo desafio para áreas mais diretamente ligadas ao uso de Tecnologias da Informação e Comunicação e, neste cenário, à Ciência da Informação cabe papel preponderante. Torna-se necessário estudar e propor melhorias em todas as fases do processo, desde o planejamento até a visualização dos dados, o que exige a proposta de um modelo de identificação e descrição destas fases, propiciando o alinhamento de esforços, recursos e resultados. Propõe-se neste trabalho um modelo de Ciclo de Vida dos Dados que leve em conta características e especificidades da Ciência da Informação, partindo-se de pesquisa realizada sobre modelos de ciclo de vida de áreas correlatas. Dentre os modelos estudados, cabe destaque o proposto pela Data Documentarion Initiative – DDI, que foi base para outros e adotado iniciativas de gestão de dados de bibliotecas que são referência mundial como a do Massachusetts Institute of Technology – MIT. O modelo proposto neste trabalho, diferencia-se dos anteriores por balizar-se nos pressupostos da Ciência da Informação, mantendo ainda um foco na vinculação com a aplicação dos conceitos teóricos. Conclui-se a premente necessidade de suscitar a reflexão e o debate sobre a definição de um modelo de ciclo de vida dos dados que seja compartilhado pela Ciência da Informação, fortalecendo a compreensão comum dos esforços e resultados das pesquisas realizadas, e que pode ter como ponto de partida a proposta de um Ciclo de Vida dos Dados, embasado nas funções preestabelecidas da área.

Palavras-Chave: Acesso a Dados. Ciclo de Vida dos Dados.

Abstract

The increasing importance of access to data has achieved in all areas presents itself as a new challenge to the most directly linked to the use of Information Technologies and Communication, in this scenario, it is up to the Information Science major role. It is necessary to study and propose improvements in all phases of the process, from planning to the visualization of data, which requires the proposal of a model of identification and description of these phases, allowing the alignment of efforts, resources and results. It is proposed in this paper a model Lifecycle Data to take into account specificities of Information Science, starting with research on models of life cycle related areas. Among the models studied, it is worth mentioning the date proposed by Documentarion Initiative - DDI, which was the basis for other initiatives adopted and data management libraries that are world reference as the Massachusetts Institute of Technology - MIT. The model proposed in this work differs from previous beacon for the presuppositions of Information Science, while still maintaining a focus on linking with the application of theoretical concepts. The conclusion is the urgent need to raise the reflection and debate on the definition of a model of the life cycle of data that is shared by the Information Science, strengthening common understanding of the efforts and results of research conducted, and that can take as its point starting a proposed Life Cycle Data, based on the functions of the predetermined area.

Keywords: Data Access. Data Life Cycle.

1 INTRODUÇÃO

Desde o princípio do registro de informações em suportes, um dos objetivos era o de propiciar acesso a dados sobre inventários, produção e comercialização, de tal forma que fosse possível manter uma percepção posterior sobre o que se possuía e acontecia. Este objetivo é cada vez mais o foco de atenção em todas as áreas de atividade e do conhecimento humano, sendo que, o grande potencial proporcionado pelas tecnologias digitais, tornou possível um novo patamar de usos e resultados na gestão e acesso a dados.

Transformar este potencial em uso concreto requer novos conhecimentos e técnicas, cobrindo todas as fases do acesso a dados, desde o planejamento sobre a coleta e geração até a visualização e neste sentido ganha importância a participação de todas as áreas do conhecimento, como por exemplo na elaboração, gestão e manutenção de recursos tecnológicos pela Ciência da Computação, na busca por aprimoramentos da análise pela Matemática, e por os todas as demais no acompanhamento e adequação de especificidades próprias de cada necessidade.

No entanto, cabe a Ciência da Informação (CI) papel preponderante na construção de novo arcabouço teórico e na definição de caminhos para que estes recursos contribuam para o atendimento das necessidades informacionais, já que cabe a esta ciência o papel de investigar o comportamento da informação, seu fluxo e os meios para o seu acesso (BORKO, 1968; CAPURRO, 2003) destacando, ainda, que a CI pode auxiliar no processo de “transformar a imensa massa de dados operacionais disponíveis diariamente em informações consistentes que permitam a tomada de decisões e agreguem valor às atividades e aos negócios” (SANTOS E VIDOTTI, 2009). Vale destacar também a relevância social desta responsabilidade da CI retomando a afirmação de que “transmitir o conhecimento para aqueles que dele necessitam é uma responsabilidade social, e essa responsabilidade social parece ser o verdadeiro fundamento da Ciência da Informação” (WERSIG & NEVELLING, 1975)

E para desempenhar esta missão, torna-se fundamental, conhecer e contribuir em todas as fases e fatores do processo de acesso a dados, o que leva a necessidade de se elaborar um modelo que sirva de base para compreensão sobre: quais são estas fases; como elas se relacionam; quais os fatores envolvidos em cada uma delas; quais os recursos disponíveis; como tirar o melhor proveito de cada uma delas, e; como é possível melhorá-las.

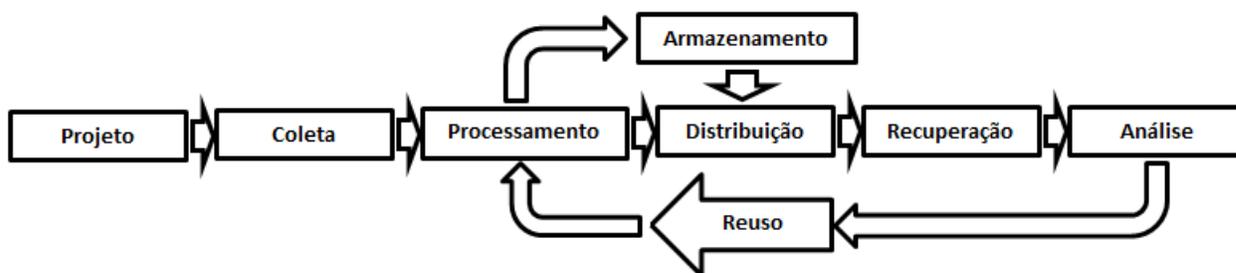
Este foco maior nos dados é relativamente recente e ainda não é possível ter uma visão completa sobre todos os aspectos envolvidos, mas neste processo de desenvolvimento é crucial uma compreensão comum de pontos chave independentemente de tecnologias, linguagens ou sistemas específicos (MÖLLER, 2013).

Neste estudo, buscou-se apresentar uma proposta de modelo de Ciclo de Vida dos Dados para a CI por meio de pesquisa sobre modelos similares e identificando características pertinentes a área.

2 CICLO DE VIDA DOS DADOS

Com o objetivo de propor um modelo para a Ciência da Informação de ciclo de vida dos dados, realizou-se pesquisa bibliográfica sobre modelos já existentes, suas características e usos, buscando identificar, características comuns e que pudessem subsidiar esta proposta e entre eles, um dos que apresentaram maior destaque foi o modelo proposto pela Data Documentation Initiative – DDI, figura 1, composto por oito fases identificadas em sequência tendo sido adotado pelo sistema de bibliotecas do Massachusetts Institute of Technology – MIT¹, e como base para outros modelos conforme descrito neste trabalho.

Figura 1 – Ciclo de Vida dos Dados – DDI



Fonte: Adaptado de DDI (2004)

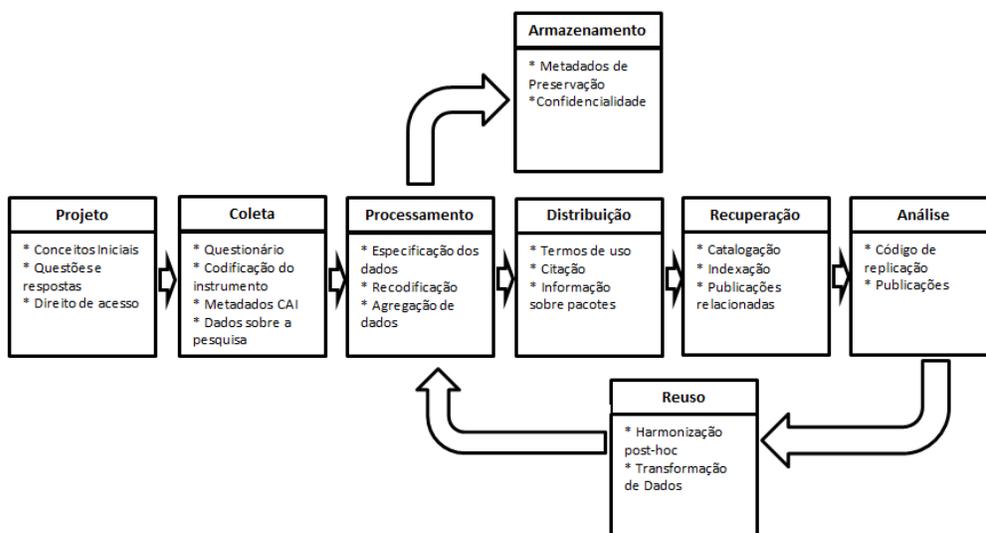
A partir da premissa de que os dados de pesquisa são ativos valiosos que devem ser administrados com cuidado em todos os momentos, (UCL, 2013) e visando ampliar a eficiência da coleta e uso destes dados o MRC Centre of Epidemiology for Child Health da University College London UCL, investiu recursos consideráveis no desenvolvimento de um sistema que permite que os envolvidos no trabalho de investigação possam trabalhar com segurança com os dados, preservando a confidencialidade dos dados pessoais e, muitas vezes sensíveis,

¹ <http://libraries.mit.edu/>

promovendo a adoção de boas práticas na coleta, processamento e arquivamento de dados, e definiu este processo como ciclo de vida dos dados de pesquisa (UCL, 2012).

O modelo (figura 2) foi baseado na proposta do DDI e apresenta como grande diferencial a inclusão, já no diagrama, das principais tarefas envolvidas em cada fase e enfoque em dados coletados por meio de entrevistas. Esta característica fica evidente na própria definição de tarefas, como na fase de coleta em que se destaca o questionário, metadados CAI (entrevista assistida por computador) e paradata que são os dados sobre como a pesquisa foi realizada. Outra característica deste modelo é sua finalidade como base para o sistema implementado.

Figura 2: Ciclo de Vida de Dados de Pesquisa – UCL



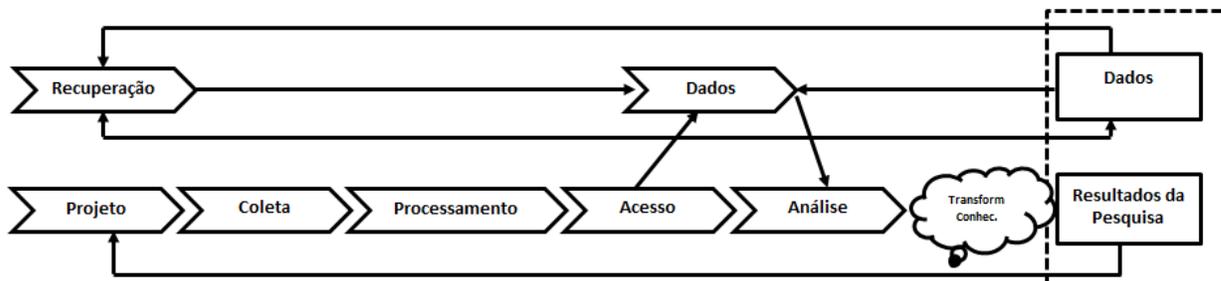
Fonte: Adaptado de UCL (2012).

Ainda com base no modelo proposto pela DDI, Humphrey (2006) elaborou uma versão de ciclo de vida dos dados (figura 3), com o diferencial de que seu escopo foi ampliado, buscando apreender o processo de criação de conhecimento empírico a partir do ponto de vista dos dados gerados. Já em 2004, Jacobs e Humphrey afirmavam que:

O arquivamento de dados é um processo, não um estado final onde os dados são simplesmente entregues a um repositório na conclusão de um estudo. Em vez disso, o arquivamento de dados deve começar no início de um projeto e incorporar uma agenda para o depósito de produtos ao longo do ciclo de vida do projeto e da criação e preservação de metadados precisa, garantindo a usabilidade do próprio dado da pesquisa. Tais práticas poderiam incorporar o arquivamento como parte do método de pesquisa. (JACOBS e HUMPHREY, 2004)

Neste modelo, Humphrey buscou ilustrar as principais considerações para o arquivamento em cada etapa do processo de criação de dados, destacando que o processo não é tão linear quanto o diagrama sugere, mas que o modelo tem sua importância como elemento balizador no desenvolvimento de um plano para abordar as considerações que entram em jogo em todas as fases do ciclo de vida dos dados (HUMPHREY, 2006).

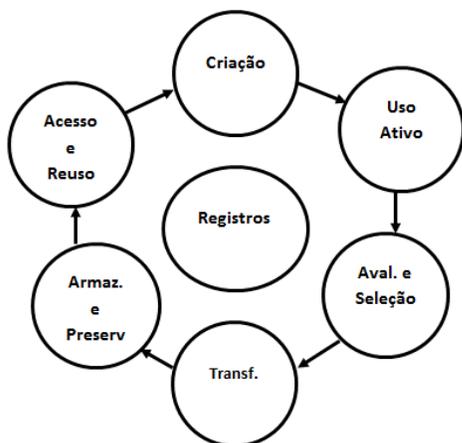
Figura 3: Modelo de Ciclo de Vida da Criação de Conhecimento de Pesquisa



Fonte: Adaptado de Humphrey (2006).

O ciclo de vida da gestão de registros pode ser acompanhada pelo diagrama proposto por Pennock (2007) identificado na figura 4, que tem em suas fases, como principal diferencial de outros modelos de ciclo de vida, a identificação da fase de uso ativo em que o registro é utilizado enquanto parte de uma base de dados e apresenta como elemento de retroalimentação a fase de acesso e reuso que pode gerar a criação de novos registros.

Figura 4: Ciclo de Vida de Dados - Pennock

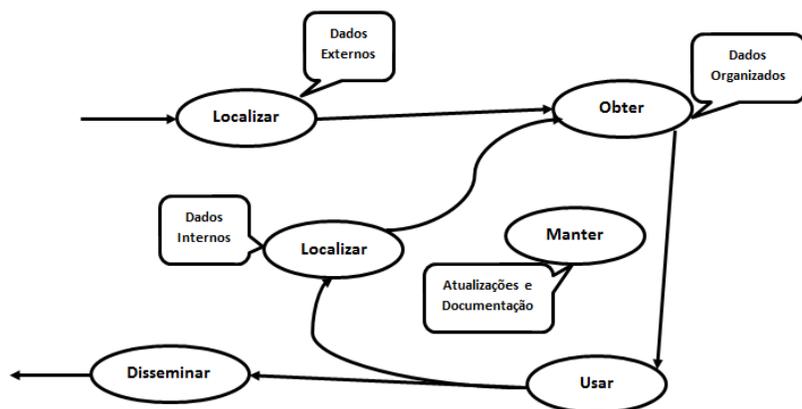


Fonte: Adaptado de Pennock (2007).

Com relação a aplicação prática da gestão do processo de acesso a dados, um exemplo de ciclo de vida de dados (figura 5) é o proposto por Ferderer (2001) e adotado por diversas

instâncias do governo americano tais como o Natural Park Service/U.S.Depto. of the Interior² e U.S.Geological Survey³, entre outros. E ainda segundo Ferderer, 2001: "Em um modelo de ciclo de vida, dados e informações são transformados em produtos e conhecimento tangíveis por um fluxo contínuo em que a saída de um processo torna-se a entrada dos outros.". Neste modelo são apresentados insumos e resultados do processo como produtos bem delineados e definidos, o que lhe confere a característica de aplicabilidade.

Figura 5: Ciclo de Vida de Dados – Componentes e recursos estratégicos



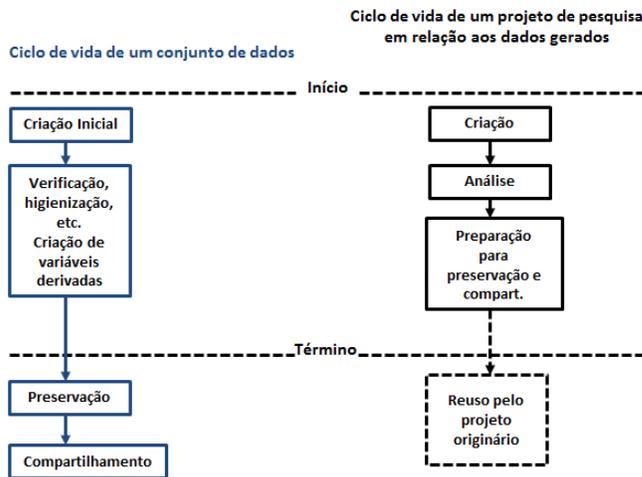
Fonte: Adaptado de Ferderer (2001).

Na mesma linha de modelos de ciclo de vida de dados que privilegiem o aspecto prático, merece ser citado o modelo proposto pela UKDA United Kingdom Data Archive e ESDS Economic and Social Data Service oferece uma visão do ciclo de vida de dados, correlacionando as fases de projeto para sua criação, proporcionando um cenário de análise para o acompanhamento de cada uma das fases.

² <http://science.nature.nps.gov/im/>

³ <http://www.usgs.gov/>

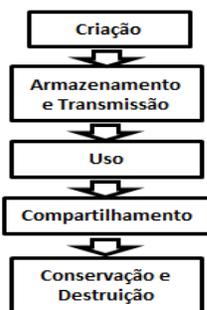
Figura 6: Ciclo de Vida de Dados comparado com o projeto que o cria.



Fonte: Adaptado de UKDA (2013)

O modelo proposto (figura 7) pela Data Management Association DAMA⁴, organização sem fins lucrativos, formada pela associação de profissionais e técnicos e dedicados a promover os conceitos e melhores práticas da Gestão da Informação e Governança de Dados. traz como contribuição ao processo a figura da operação de descarte dos dados, sendo tratada como última fase do ciclo, sendo realizada em conjunto com as ações de preservação, identificada como “*retention and destruction*”.

Figura 7: Ciclo de Vida dos Dados – DAMA

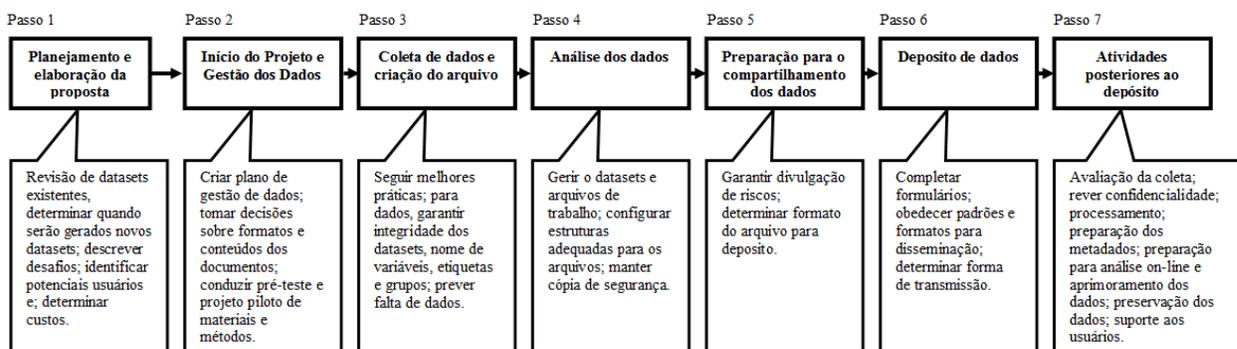


Fonte: <http://www.dama.org> (tradução do autor)

Já no modelo proposto pela Data Sharing for Demographic Research DSDR (figura 8) as fases congregam mais de uma tarefa e apresentam no diagrama estes objetivos, contribuindo para o entendimento do processo como um todo.

Figura 8: Ciclo de Vida de Dados – DSDR

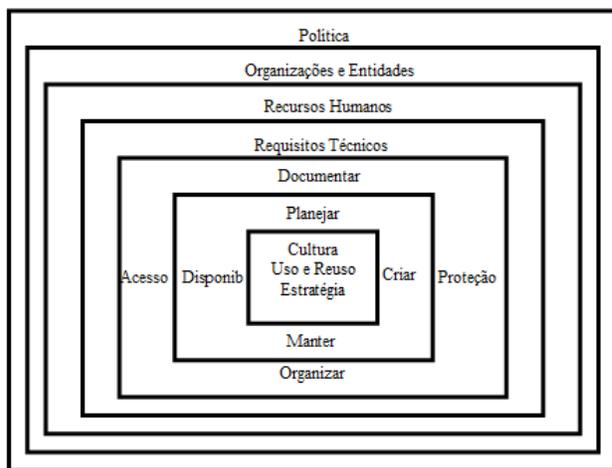
⁴ <http://www.dama.org.br/>



Fonte: <http://www.icpsr.umich.edu/icpsrweb/DSDR/>

O modelo proposto pela Interagency Working Group on Digital Data (IWGDD) extrapola o fluxo informacional em si e incorpora elementos contextuais como a própria organização, requisitos técnicos, recursos humanos e perfis profissionais necessários e políticas de informação, figura 9. Neste modelo vê-se a fase de planejamento quando se determina quais dados precisam ser criados ou recolhidos para apoiar uma agenda de pesquisa ou uma missão/função, identifica-se e avalia-se as fontes existentes de dados necessários além de padrões de dados e formato de metadados, definindo-se ainda, ações e responsabilidades pela gestão dos dados durante seu ciclo de vida. Na fase de criação, são produzidos ou adquiridos dados para os fins previstos e ainda são criados os recursos de apoio para o acesso aos dados de acordo com as finalidades definidas. Na fase de manutenção os dados são organizados e armazenados para atender as necessidades especificadas no planejamento, garantindo inclusive sua preservação. Na fase de disposição são estabelecidas estratégias de finalização da base, com planejamento para transferência dos dados para outra base ou descarte (IWGDD, 2009).

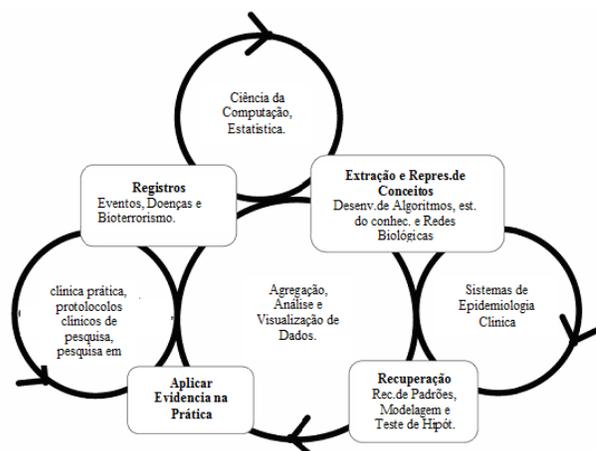
Figura 9: Modelo de Ciclo de Vida de Dados Digitais – IWGDD



Fonte: Adaptado de IWGDD (2009).

Modelos que representam necessidades específicas, apesar de direcionados por características próprias, podem contribuir com aspectos importantes sobre os processos envolvidos no acesso a dados, como é o caso do modelo (figura 10) utilizado pela Vanderbilt University Medical Center⁵ – VUMC, com a preocupação de descrever o processo de incorporação de resultados de pesquisa a prática e a partir dela a retroalimentação a novas pesquisas. Identificado como “*Translational Research*” o diagrama é composto por quatro ciclos que apresentam a percepção do acesso aos dados em uma sequência em que os mesmos são coletados e consumidos e neste processo são gerados insumos para um novo processo.

Figura 10: Ciclo de Vida dos Dados e a Pesquisa Translacional



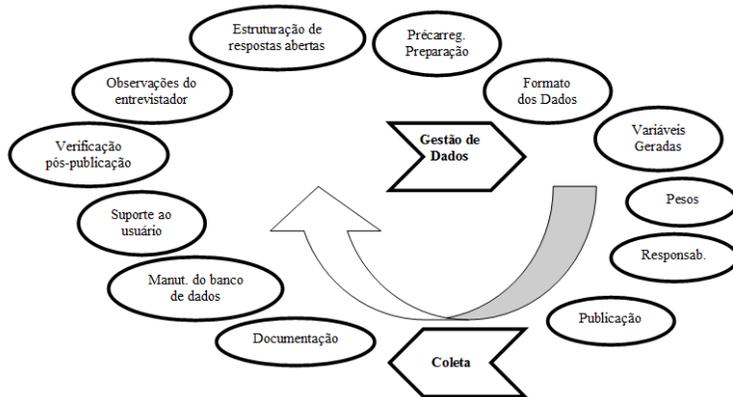
Fonte: VUMC (2005), tradução do autor.

A partir de um projeto que coleta grandes quantidades de dados como o Survey of Health, Ageing and Retirement in Europe (SHARE) surgem muitos requisitos no processo de acesso a estes dados e o modelo apresentado na figura 11 descreve uma sequência circular e linear de fases, acrescentando ainda as tarefas e atores envolvidos em cada uma. Este projeto ainda apresenta a peculiaridade de tratar de dados públicos sobre saúde e que são por natureza sensíveis. O grande volume de dados brutos requer grandes alterações para que se tornem acessíveis, principalmente na formatação e disponibilização de variáveis geradas. Por se tratar de dados sensíveis o tratamento dos dados passa também por um cuidado especial na codificação das

⁵ <https://ncs.mc.vanderbilt.edu>

respostas obtidas. Os autores destacam ainda que o processo não é unidirecional, mas sim um processo que inclui *loops de feedback*, ou seja, determinadas correções nos dados só podem ser detectadas a partir de novas informações.

Figura 11: Tarefas e atores envolvidos no gerenciamento da base de dados da SHARE



Fonte: Adaptado de Hunkler et al. (2011).

A área ambiental tem crescido em relevância e em foco e representa uma fronteira de pesquisa sobre acesso a dados, principalmente em função do grande volume de dados gerados, volume este que tende a crescer de forma exponencial em função da disseminação de sensores e coletores de dados cada vez mais acessíveis. A Data Observation Network for Earth (DataONE) é um exemplo desta nova ciência ambiental inovadora, buscando atuar através de estruturas distribuídas e sustentáveis que atendam às necessidades da ciência e da sociedade de acesso aberto, persistente, robusto e seguro de dados observacionais da Terra, bem descritos e de fácil acesso. Para atender esta demanda, a DataONE propõem um modelo de ciclo de vida de dados (figura 12) que possa nortear o planejamento e realização da gestão destes dados. (DataONE, 2013)

Figura 12: Ciclo de Vida dos Dados – DataONE



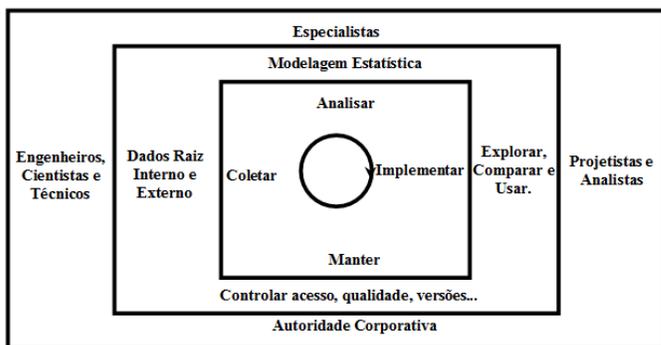
Fonte: DataONE (2013) tradução do autor.

Outro modelo (figura 13) que representa uma necessidade específica é o proposto pelo Material Data Management Consortium - MDMC que identifica o objeto e os responsáveis por

cada fase. O consórcio MDMC é resultado da colaboração entre a ASM International⁶, Granta⁷, NASA Glenn Research Center e muitas outras empresas internacionais do setor aeroespacial, de defesa e de energia e tem como missão identificar melhores práticas de gestão, análise e uso de informações sobre materiais, vinculando suas iniciativas a aplicação (MDMC, 2013) e daí a importância deste diagrama nesta pesquisa.

Ainda considerando a necessidade de descrever os processos envolvidos no acesso a dados, vale destacar o conceito de curadoria de dados. O Digital Curation Centre – DCC, que tem como foco a capacitação para a gestão de dados de pesquisa em toda comunidade de pesquisa de ensino superior do Reino Unido, propõe um modelo de ciclo de vida de curadoria de dados (figura 14) que apresenta características que contribuíram com o modelo CVD-CI proposto neste trabalho, em especial a distinção de ações que ocorrem durante todo o ciclo de vida tais como: Descrição e Representação da Informação, que consistem em atribuir metadados administrativos, descritivos, técnicos, estruturais e de preservação, utilizando padrões adequados, para assegurar a descrição e o controle adequados a longo prazo, e ainda, coletar e atribuir informações de representação necessários para compreender e processar tanto o material digital como os metadados associados (DCC, 2013).

Figura 13: Ciclo de Vida de Dados sobre Materiais



Fonte: Adaptado de MDMC (2013).

Outra ação que também ocorre em todo o ciclo é: Planejamento de Preservação que inclui planos para a gestão e administração de todas as ações do ciclo de vida de curadoria. Da mesma forma tem-se a Acompanhamento da Comunidade e Participação que visa manter uma vigilância sobre as atividades comunitárias apropriadas e participar no desenvolvimento de normas comuns, ferramentas e software adequado, e completando as ações neste nível tem-se a Curadoria e

⁶ <http://www.asminternation.org>

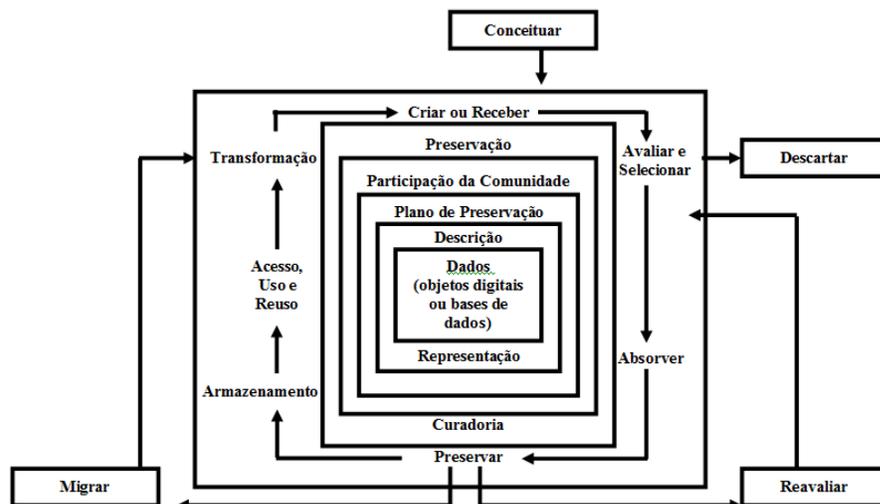
⁷ <http://www.granta.com>

Preservação que visam realizar a gestão e ações administrativas planejadas para promover a curadoria e preservação ao longo do ciclo de vida da curadoria. (DCC, 2013)

Assim, como no CVD-CI, neste modelo tem-se as ações seqüências, que neste caso estão divididas em: conceituação; criação; avaliação e seleção; inserção; preservação; armazenamento; e transformação. Na fase de inserção ocorre a transferência do dado para um arquivo, um repositório, um data center ou outro suporte e são verificados os aspectos legais e da aderência a política de informação adotada. São citadas ainda ações ocasionais como o descarte de dados que não sejam mais úteis, reavaliações e migração de dados para outros formatos (DCC, 2013).

Não seria possível tratar o ciclo de vida dos dados sem levar em consideração a importância da contextualização dos dados armazenadas, sejam eles administrativos, descritivos, técnicos, estruturais ou de preservação, a insuficiente carga semântica, própria dos dados, torna o uso de metadados um fator chave de sucesso.

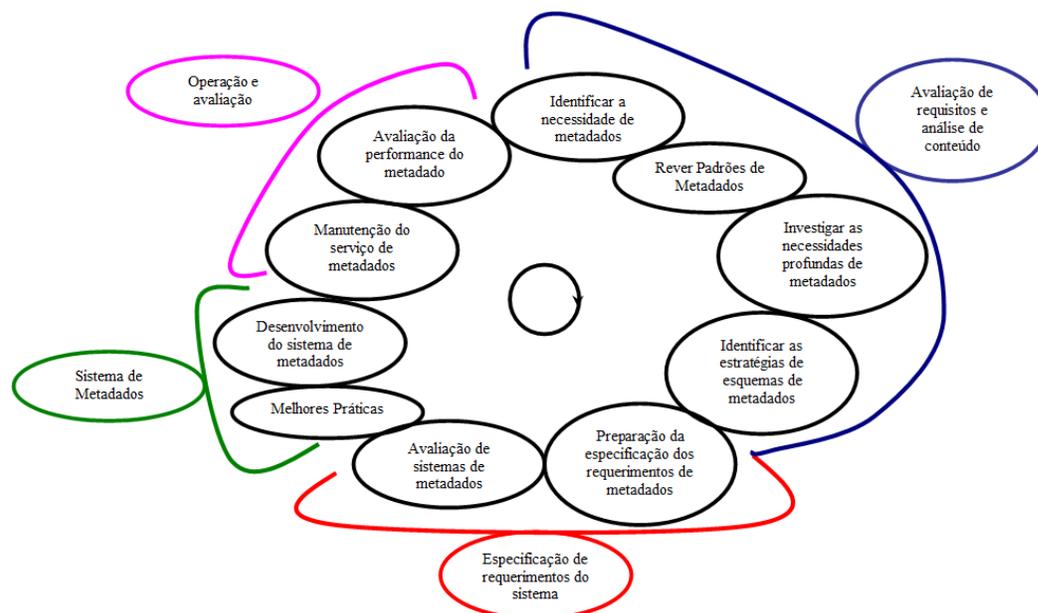
Figura 14: Ciclo de Vida de Curadoria dos Dados



Fonte: DCC, 2013.

Há de se considerar ainda a necessidade de padrões adequados para assegurar a descrição e o controle ao longo do tempo. Assim, acrescenta-se neste estudo o ciclo de vida de metadados proposto por Che et al., 2003, sem que seja possível, no entanto, um maior aprofundamento na análise do mesmo neste texto por estar fora do escopo proposto.

Figura 15: Ciclo de Vida de Metadados



Fonte: Adaptado de CHEN et al (2003).

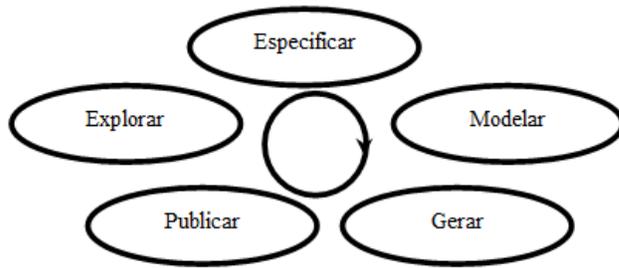
Outra tendência relacionada ao acesso aos dados e a questão semântica, refere-se a questão dos Dados Ligados (*Linked Data*), ou seja dados agregados por relacionamentos a outros dados e que assim incorporam carga semântica neste processo de relacionamento. Especial atenção tem sido dada a questão dos *Linked Open Data*, que acrescentam a característica de dado aberto aos dados ligados, o que os tornam ainda mais relevantes.

Complementando o tema de Linked Open Data, apresenta-se (figura 16) um modelo de ciclo de vida simplificado proposto por VILLAZÓN-TERRAZAS at al (2009) que colabora no processo de compreensão das diferenças nos ciclos de vida de dados ordinários e de dados ligados. Neste modelo tem-se as fases de Especificação em que ocorre a identificação e análise de fontes de dados governamentais, o projeto de Uniform Resource Identifier – URI e ainda a definição da licença. Na fase de Modelagem tem-se o compartilhamento da compreensão do domínio de interesse.

Já na fase de Geração ocorre a transformação, quando as fontes de dados selecionados na atividade de especificação são transformadas em RDF de acordo com o vocabulário criado na atividade de modelagem. Ainda nesta fase ocorre a “limpeza” dos dados e a instanciação das ligações entre os dados.

Na fase de Publicação ocorre a divulgação e disponibilização dos dados, dos metadados e dos recursos de recuperação. (VILLAZÓN-TERRAZAS at al, 2009)

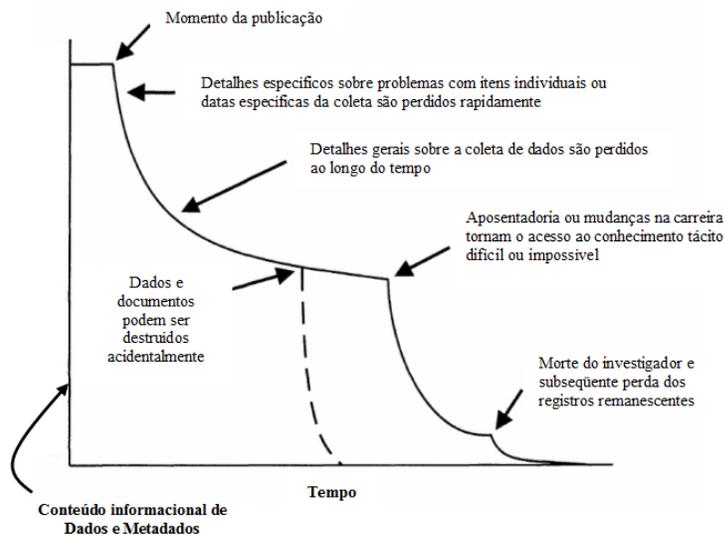
Figura 16: Ciclo de Vida de Linked Open Data



Fonte: Adaptado de VILLAZÓN-TERRAZAS et al (2009).

Com relação a fase específica de descarte proposta no modelo CVD-CI apresentado neste texto, apresenta-se o gráfico de Entropia de Dados (figura 17) proposto por Michener et al. (1997) e que explicita situações de degradação do conteúdo de dados e dos metadados em função do tempo contanto inclusive com exemplo da degradação normal em conteúdo de informação associados aos dados e metadados ao longo do tempo e acidentes ou alterações na tecnologia de armazenamento (linha tracejada) que podem eliminar o acesso ao restante dos dados brutos e dos metadados, em qualquer momento.

Figura 17: Entropia de Dados



Fonte: Adaptado de Michener et al. (1997)

Para concluir os ciclos de dados pesquisados, apresenta-se o ciclo de vida de dados e publicações (figura 18) e que é bastante citado, principalmente na Ciência da Informação, descrevendo as fases envolvidas no processo de publicação de produções científicas. No entanto

este modelo não atende as necessidades específicas oriundas do processo de acesso a dados de modo mais geral.

Figura 18: Ciclo de Vida de Dados e Publicações



Fonte: Adaptado de Gold (2007).

3 FASES E OBJETIVOS

Faz-se necessário diferenciar fases e objetivos na elaboração do Ciclo de Vida dos Dados e no caso da Ciência da Informação identificou-se três fases no processo de acesso a dados: Coleta, Armazenamento e Recuperação, enquanto que alguns objetivos permeiam estas fases.

Para compreensão desta distinção vamos ao exemplo da preservação, que é um objetivo que, se em uma primeira análise parece ser uma fase final do ciclo, apresenta-se como um objetivo que deve ser considerado desde o momento da coleta, em que é levado em consideração não somente no planejamento e na elaboração do formato e dos metadados, mas também na própria preocupação com a preservação dos metadados em si e dos aspectos relacionados a sua interpretação. O mesmo acontece nas fases de armazenamento, com os aspectos relativos ao suporte e ainda na fase de recuperação com as preocupações sobre os recursos de consulta e de visualização. Assim, é um fator que ocorre durante todo o processo e não em uma fase específica. O mesmo raciocínio se aplica a objetivos como a qualidade, privacidade e direitos autorais.

Assim, ao analisar os exemplos de modelos de ciclo de vida de dados, percebe-se que existem aqueles que apresentam somente fases⁸ interligadas de forma linear e outros que apresentam fases permeadas por objetivos⁹ e, considerando o estudo do comportamento da informação, seu fluxo e os meios para o seu acesso como objetivos da CI, propõe-se para efeito de compreensão do processo de acesso a dados as fases de: coleta, armazenamento, recuperação e

⁸ (DDI, Humphrey, SHARE, DataONE, Pennock, Ferderer, Dama, ICPSR, UKDA, DSDR, VUMC, LOD, MDMC, Villazón-Terrazas, CHEN)

⁹ (DCC, IWGDD)

descarte, e que são permeadas pelos objetivos de: privacidade, qualidade, direitos autorais, integração, disseminação e preservação.

Ao se comparar as fases propostas para o modelo de ciclo de vida dos dados para Ciência da Informação (CVD-CI) com os modelos identificados na pesquisa (quadro 1), e os objetivos no CVD-CI e sua correlação com os demais modelos (quadro 2), percebe-se que :

- fases do CVD-CI agregam mais de uma fase em outros modelos;
- algumas fases dos modelos estudados foram consideradas no CVD-CI como objetivos;
- no caso do diagrama de Pennock o uso ativo e a avaliação e seleção foram identificados como parte da fase de coleta por se entender que o registro em questão ainda não esta sendo considerado como base para disponibilização, já que se encontra em fase anterior a transferência e armazenamento, assim, apesar de já poder estar em um suporte digital, encontra-se em fase anterior a composição da base a ser utilizada em processos de recuperação;

- na comparação com o modelo DSDR, foram consideradas as descrições encontradas em cada uma das fases, já que as mesmas congregam um conjunto grande de tarefas e, assim, partes de uma fase foram vinculadas a fases e objetivos do CVD-CI e como no caso da fase “*archival activities*” houve vinculação a mais de uma fase e a mais de um objetivo;

Nem todos os modelos de ciclo de vida encontrados foram incluídos nos quadros comparativos por serem sintéticos demais, como nos casos do modelo proposto em MDMC e que apresenta somente as fases: analisar, implementar, manter e capturar, e também do modelo proposto por Ferderer que é composto pelas fases: encontrar, obter, utilizar e disponibilizar. A mesma situação ocorre com o modelo proposto pela IWGDD que é composta pelas fases: planejar, criar, manter e disposição.

Quadro 1. Comparativo de fases nos modelos de ciclo de vida de dados.

CVD-CI	DDI	DataONE	DCC	Pennock	DAMA	DSDR
Coleta	conceito coleta	planejar coletar descrever	conceituar criar receber representar avaliar selecionar	criação uso_ativo avaliação seleção	criação	proposta planejamento início gestão coleta criação análises
Armazenamento	processamento arquivo		inserir armazenar transformar reavaliar migrar	transferência armazenamento	armazenamento transmissão	depositar arquivamento
Recuperação	recuperação análise reuso	recuperar analisar	acesso	acesso Reuso	uso	arquivamento

Descarte			dispor		destruição	
----------	--	--	--------	--	------------	--

Fonte: autor.

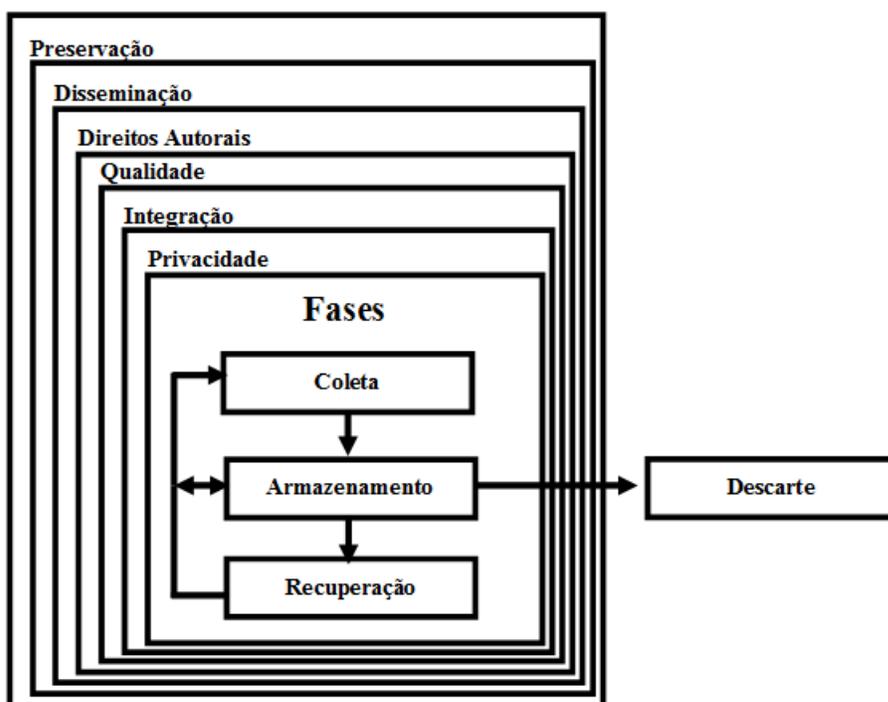
Quadro 2. Correlação entre os objetivos do CVD-CI e os modelos pesquisados.

CVD-CI	DDI	DataONE	DCC	Pennock	DAMA	DSDR
Privacidade						planejamento
Qualidade		qualificar				arquivamento
Direitos Autorais						
Integração		integrar				
Disseminação	distribuição				compartilhamento	compartilhar
Preservação		preservar	preservar	preservação	retenção	arquivamento

Fonte: autor.

Os próprios quadros comparativos (1 e 2) auxiliam na interpretação das fases e objetivos propostos no CVD-CI, e o diagrama com as fases e objetivos pode ser representado conforme figura 19.

Figura 19. Ciclo de Vida dos Dados para Ciência da Informação – (CVD-CI)



Fonte: autor.

No processo descrito no CVD-CI (figura 19) a fase de coleta permite que seja iniciada a fase de armazenamento que por sua vez propicia a execução da fase de recuperação e ainda pode gerar novos dados retomando ações da fase de coleta.

Uma vez em execução atividades da fase de recuperação, tem-se mais uma vez a possibilidade de geração de novos dados originados nesta fase o que retroalimenta o ciclo e retoma ações da fase de coleta para situações novas ou diretamente a fase de armazenamento para dados que já são esperados como resultado das ações ocorridas na fase de recuperação.

Alcançados os objetivos previamente estabelecidos e ou alcançados os limites de tempo previstos para os dados, passa-se a fase de descarte, que dependendo da situação irá gerar dados, seja por transferência para outras bases seja para efeito de preservação de informações selecionadas da base a ser descartada, originando assim o início de atividades, ora da fase coleta, ora da fase de armazenamento.

Na fase de coleta têm-se as atividades vinculadas a definição inicial dos dados a serem utilizados, seja na elaboração do planejamento de como serão obtidos, filtrados e organizados, identificando-se a estrutura, formato e meios de descrição que será utilizado. Nesta fase o dado deve ainda ser devidamente descrito em metadados, avaliados e selecionados. Nos casos em que a fonte seja oriunda de registros de uso corrente de uma base de dados operacional, tem-se ainda nesta fase a identificação das ações pertinentes a formação da nova base e que ainda ocorram na base origem.

A fase armazenamento representa as atividades relacionadas ao processamento, transformação, inserção, modificação, migração, transmissão e toda e qualquer ação que vise a persistência de dados em um suporte digital.

A recuperação é a fase em que o acesso aos dados se concretiza, ocorrendo portanto, as atividades ligadas a consulta e visualização, com as etapas posteriores a obtenção do dado: estruturação, filtro, tratamento, representação, refinamento e interatividade. Vale destacar que a visualização dos dados, apesar de tarefa fundamental na etapa de recuperação de dados não foi citada por nenhum dos modelos de representação do ciclo de vida de dados.

4 CONCLUSÕES

Com o aumento da demanda por acesso a dados, faz-se necessário o aumento de recursos de coleta, armazenamento e recuperação deste importante insumo a todas as atividades e a Ciência da Informação tem papel de destaque nesta tarefa.

A compreensão comum dos aspectos envolvidos no ciclo de vida dos dados é fator chave de sucesso no desenvolvimento de pesquisas e de novas tecnologias e os modelos de ciclo de vida

existentes não atendem as características encontradas na Ciência da Informação o que suscita a necessidade de um modelo próprio da área e que possa, inclusive ter um escopo tanto de apoio teórico como na aplicação das tecnologias.

Um dos principais objetivos deste trabalho é suscitar a reflexão e o debate sobre a definição de um modelo de ciclo de vida dos dados que seja compartilhado pela Ciência da Informação, fortalecendo a compreensão comum dos esforços e resultados das pesquisas realizadas, e lança como ponto de partida a proposta de um Ciclo de Vida dos Dados, embasado nas funções preestabelecidas da área e que, se espera, seja mais um elemento de fomento a pesquisas, na Ciência da Informação sobre o tema “Acesso a Dados”.

REFERÊNCIAS

BORKO, H. **Information science**: what is it? *American Documentation*, v.19, n.1, p.3-5, Jan. 1968.

BRASIL. Congresso. **Lei n. 9.610**, de 19 de fevereiro de 1998. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/19610.htm>. Acessado em: <10 jun 2013>

CAPURRO, R. **Epistemologia e ciência da informação**. Anais do V Encontro Nacional de Pesquisa em Ciência da Informação - ENANCIB. Belo Horizonte, 2003. (cdrom).

CHEN, Ya-Ning, CHEN, Shu-Jiun, LIN, Simn C. **A metadata lifecycle model for digital libraries**: methodology and application for an evidence-based approach to library research. In *World Library and Information Congress: 69th IFLA General Conference and Council*, ago. 2003. Disponível em <http://archive.ifla.org/IV/ifla69/papers/141e-Chen_Cheng_Lin.pdf>. Acessado em: <10 mai 2013>.

DataONE Data Observation Network for Earth. **Best Practices**. Disponível em: <<http://www.dataone.org/best-practices>>. Acessado em: <10 mai 2013>

DCC Digital Curation Center. **Curation Lifecycle Model**. Disponível em: <<http://www.dcc.ac.uk/resources/curation-lifecycle-model>>. Acessado em: <12 jan 2013>

DDI - Data Documentation Initiative: Structural Reform Group. **DDI Version 3.0 Conceptual Model**. DDI Alliance. 2004. Disponível em: <<http://libraries.mit.edu/guides/subjects/data-management/cycle.html>>. Acessado em: <02 dez 2012>

FERDERER, David A. **A data management life-cycle**. USGS Fact Sheet: 163-00. 2001. Disponível em: <<http://www.usgs.gov/>>. Acessado em: <10 mai 2013>.

GOLD, Anna. **Cyberinfrastructure, Data, and Libraries**, Part 1: A Cyberinfrastructure Primer for Librarians. D-Lib Magazine. Volume 13 Number 9/10. ISSN 1082-9873. September/October 2007. Disponível em: <<http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1015>>. Acessado em: <12 jan 2013>

HUMPREY, Charles. **e-Science and the Life Cycle of Research**. 2006. Disponível em: <<http://datalib.library.ualberta.ca/~humphrey/lifecycle-science060308.doc>>. Acessado em: <15 mai 2013>

HUNKLER, Christian, KNEIP, Thorsten, KORBMACHER, Julie, STUCK, Stephanie, ZUBER, Sabrina. **Glimpsing into the Blackbox: Data Managing and Cleaning Processes** In. Schröder, M. (ed.), 2011. Retrospective Data Collection in the Survey of Health, Ageing and Retirement in Europe. SHARELIFE Methodology. MEA, Mannheim.

ICPSR Inter-university Consortium for Political and Social Research. **Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle**. 5 ed. Ann Arbor, MI, 2012. Disponível em: <<http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf>>

IWGDD - Interagency Working Group on Digital Data. **Harnessing the Power of Digital Data for Science and Society**. jan,2009. Disponível em: <http://www.nitrd.gov/About/Harnessing_Power_Web.pdf>. Acessado em: <10 mai 2013>

JACOBS, James A., HUMPHREY, Charles. **Preserving Research Data**. Communications of the ACM. 47(9): 27-29. 2004. Disponível em: <http://3stages.org/jj/w/preserving_research_data.html>. Acessado em: <15 mai 2013>

LYNCH, Clifford. **Big data: How do your data grow?** Nature 455, 28-29 (4 September 2008) | doi:10.1038/455028a; Publicado online 3 September 2008. Disponível em: <<http://www.nature.com/nature/journal/v455/n7209/full/455028a.html>>. Acessado em: <20 jan 2013>.

MDMC Material Data Management Consortium. **The Materials Data Lifecycle**. Disponível em: <<http://www.mdmc.net/pages/lifecycle.htm>>. Acessado em: <25 jan 2013>.

MICHENER, William K.; BRUNT, James W.; HELLY, John J.; KIRCHNER, Thomas B.; STAFFORD, Susan G. **Nongeospatial Metadata for the Ecological Sciences**. Ecological Applications, Vol. 7, No. 1. (Feb., 1997), pp. 330-342. Disponível em: <<http://links.jstor>>.

org/sici?sici=1051-0761%28199702%297%3A1%3C330%3ANMFTES%3E2.0.CO%3B2-W>. Acessado em: <10 mai 2013>.

MÖLLER, Knud. **Lifecycle models of data-centric systems and domains**: the abstract data lifecycle model. Semantic Web Journal. Vol. 4, Num. 1, IOS Press. 2013. DOI:10.3233/SW-2012-0060. Disponível em: <<http://www.semantic-web-journal.net>>. Acessado em: <15/07/2013>

PENNOCK, Maureen. Digital Curation: A Life-Cycle Approach to Managing and Preserving Usable Digital Information. Library & Archives, Janeiro 2007. Disponível em: <http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf>.

SANTOS, P. L. A. da C.; VIDOTTI, S. A. B. G. **Perspectivismo e Tecnologias de Informação e Comunicação: acréscimos à Ciência da Informação?** DataGramaZero, v. 10, n. 3, jun., 2009. Disponível em: <http://www.dgz.org.br/jun09/Art_02.htm>. Acesso em: <20 jun 2012>.

UCL - University College London. MRC CENTRE OF EPIDEMIOLOGY FOR CHILD HEALTH. **Data management**. Disponível em: <<http://www.ucl.ac.uk/ich/research-ich/mrc-cech/data>>. Acessado em: <15 mar 2013>

UKDA United Kingdom Data Archive **Guide to Good Practice**: Data Management (2005), p. 8. Disponível em: <<http://www.esds.ac.uk>>. Acessado em: <03 jan 2013>

VILLAZÓN-TERRAZAS, Boris, VILCHES, Luis M., CORCHO, Oscar, GÓMEZ-PÉREZ Asunción. Methodological Guidelines for Publishing Government Linked Data. Facultad de Informática, Universidad Politécnica de Madrid. Campus de Montegancedo, Madrid. Setembro, 2011. Disponível em: <<http://www.oeg-upm.net>>

VUMC Vanderbilt University Medical Center **Strategic Plan for VUMC Informatics**. 2005 Disponível em: <https://ncs.mc.vanderbilt.edu/Data/NonSecure/IC_Strategic_Plan_9-12-05.pdf>. Acessado em: <13 jan 2013>

WERSIG, G., NEVELING, U. The phenomena of interest to information science. Information Scientist, v.9, p. 127-140, 1975.